SNS COLLEGE OF TECHNOLOGY

Coimbatore-35

An Autonomous Institution

Accredited by NBA – AICTE and Accredited by NAAC – UGC with 'A+' Grade

Approved by AICTE, New Delhi & Affiliated to Anna University, Chennai

DEPARTMENT OF MECHATRONICS ENGINEERING

19MCT401-DATA ANALYTICS IN AUTOMATION SYSTEM

INTRODUCTION

- Hadoop is an Apache open source framework written in java that allows distributed processing of large datasets across clusters of computers using simple programming models.

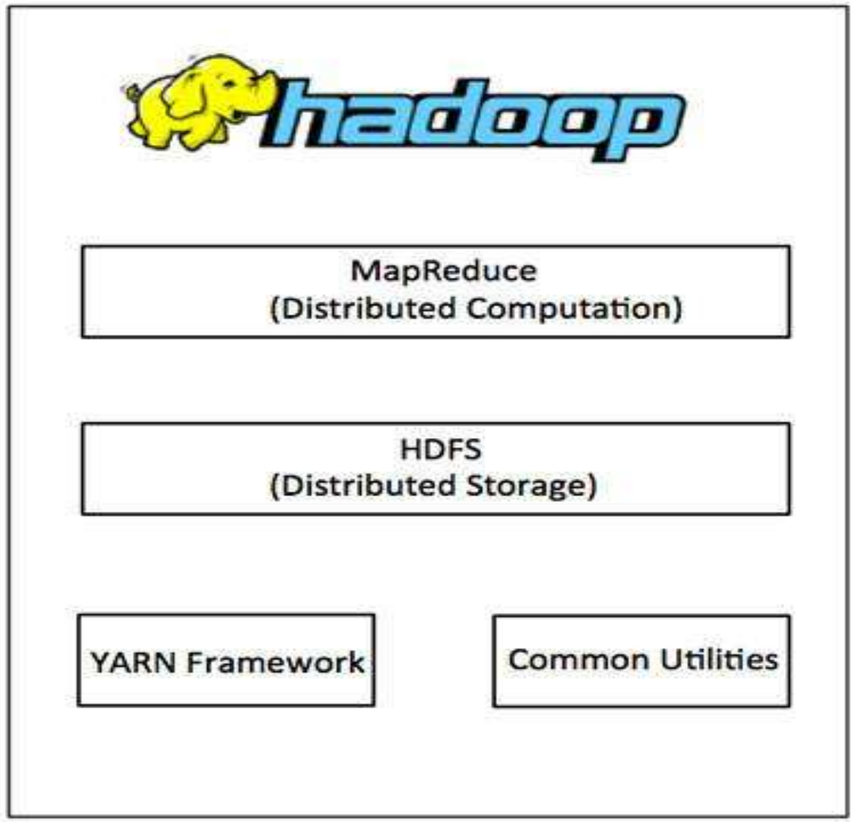
- The Hadoop framework application works in an environment that provides distributed *storage* and *computation* across clusters of computers.

- Hadoop is designed to scale up from single server to thousands of machines, each offering local computation and storage.

ARCHITECTURE OF HADOOP

Hadoop has two major layers namely ;

- Processing/Computation layer (MapReduce)
- Storage layer (Hadoop Distributed File System)



ADVANTAGES OF HADOOP

- Hadoop framework allows the user to quickly write and test distributed systems. It is efficient, and it automatic distributes the data and work across the machines and in turn, utilizes the underlying parallelism of the CPU cores.

- Hadoop does not rely on hardware to provide fault-tolerance and high availability (FTHA), rather Hadoop library itself has been designed to detect and handle failures at the application layer.

- Servers can be added or removed from the cluster dynamically and Hadoop continues to operate without interruption.

HADOOP SCALABILITY

The primary benefit of Hadoop is its Scalabillity.one can easily scale the cluster by adding more nodes.

There are two types of Scalabillity in Hadoop

- Vertical Scalabillity
- Horizontal Scalabillity

VERTICAL SCALABILITY

- It is also referred as “scale up”. In vertical scaling, you can increase the hardware capacity of the individual machine.
- In other words, you can add more RAM or CPU to your existing system to make it more robust and powerful.

HORIZONTAL SCALABILLITY

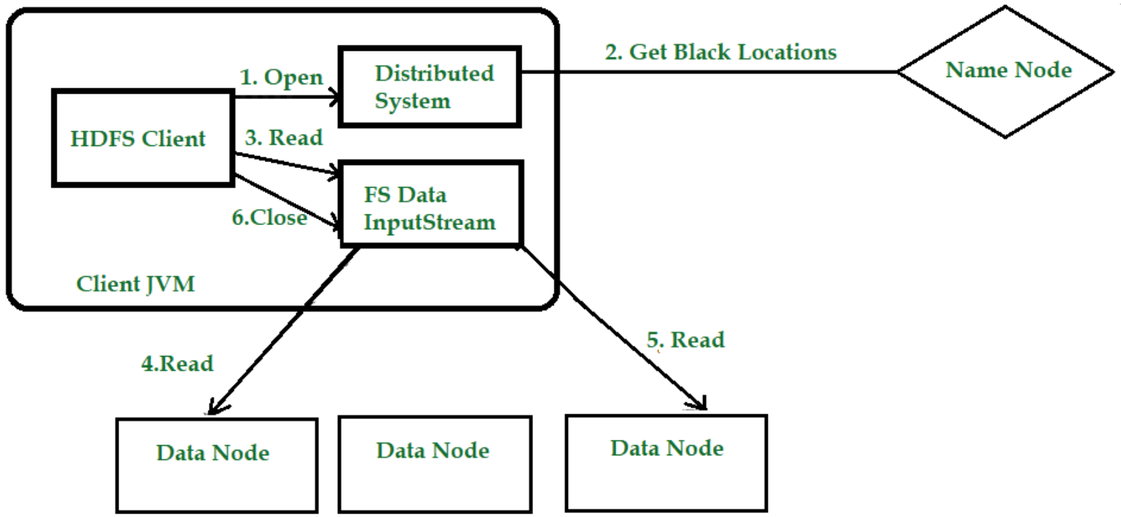
- It is also referred as “scale out” is basically the addition of more machines or setting up the cluster.

- In horizontal scaling instead of increasing hardware capacity of individual machine you add more nodes to existing cluster and most importantly, you can add more machines without stopping the system.

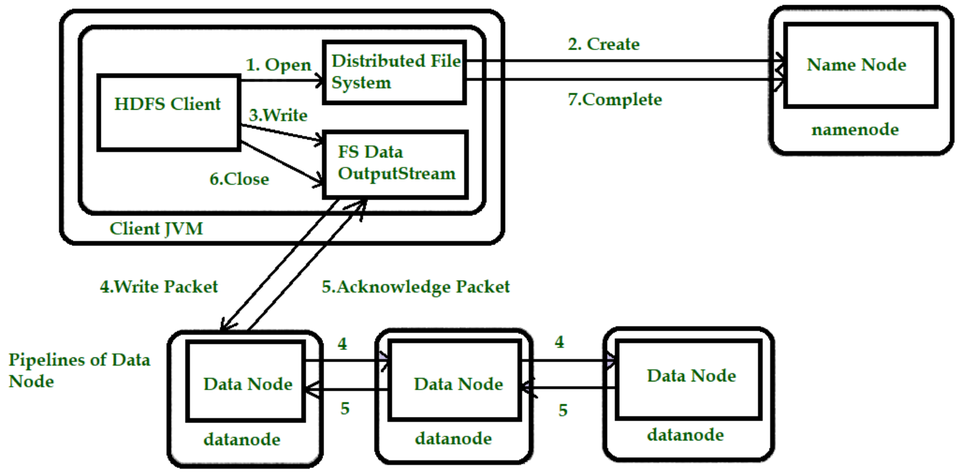
- Therefore we don’t have any downtime or green zone, nothing of such sort while scaling out. So at last to meet your requirements you will have more machines working in parallel.

ANATOMY OF HADOOP

ANATOMY OF FILE READ IN HDFS



ANATOMY OF FILE WRITE IN HDFS




Thank You